So you want to share your Data? Collaborative Policy Writing for the UMass Amherst Data Repository

Erin Jerome, Open Access & Institutional Repository Librarian, University of Massachusetts Amherst

SMIRC 2020

I.
Launching the Data Repository

Background Information*

ScholarWorks@UMassAmherst: our IR, runs on Digital Commons

May 2016: ScholarWorks contained one data file that accompanied a library-funded open access article

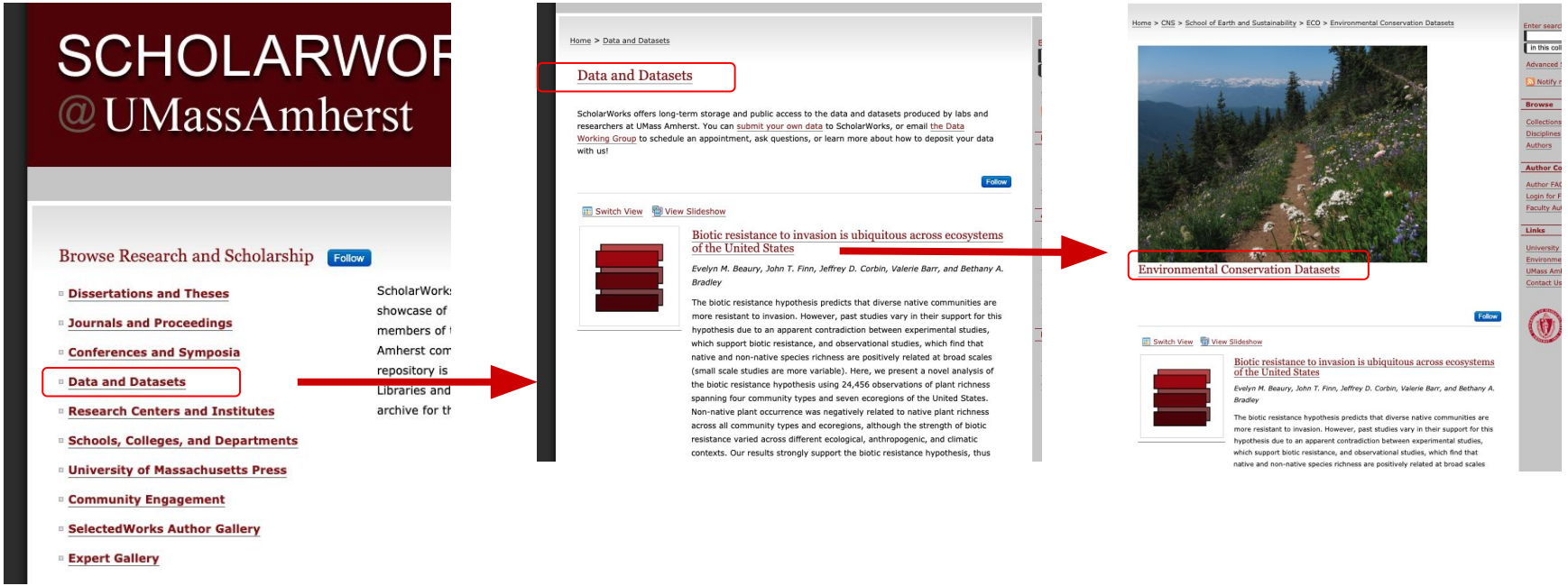
May 2016 - September 2017: Five datasets uploaded for a faculty member in Environmental Conservation whose publisher required accessible data with DOIs.

2017 - UMass Amherst Libraries hire a Data Services Librarian

Data Working Group: Cross-disciplinary team led by the Data Services Librarian

*Part of this project entailed knowingly obliterating all statistics from prior to October 2017

GOAL: Create a central Data repository



Making the Goal a Reality

Created the centralized Data and Datasets structure

Deleted and recreated Environmental Conservation Datasets using the same publication structure as Data and Datasets*

Uploaded the six Environmental Conservation datasets back to the departmental level collection to avoid broken links

*Here's where we lost those valuable stats

First Customer: Designing Sustainable Landscapes

Wanted to move all of their content from a university-hosted site to ScholarWorks

Wanted a landing page for the collection with a lot of customization

Both the Data Services Librarian and OA & IR Librarian were on extended leaves

bepress was migrating storage to the cloud - lots of timeouts and outages!

First Customer: Designing Sustainable Landscapes

The Designing Sustainable Landscapes (DSL) Project

- Ancillary Products
- Designing Sustainable Landscapes Project Technical Documents
- Ecological Impact Metrics
- Ecological Integrity Metrics
- Ecological Settings

Data Products

Ecological settings

The ecological settings products include a broad suite of static as well as dynamic abiotic and biotic variables representing the natural and anthropogenic environment at each location (cell). Static variables are those that do not change over time (e.g., elevation, incident solar radiation). Dynamic settings are available for 2010 and 2080; static settings are available for 2010. Dynamic variables are those that change over time in response to succession and the drivers (e.g., growing season degree days, traffic rate). Most of the settings variables are continuous and thus represent landscape heterogeneity as continuous (e.g., slope, biomass), although some are categorical and thus represent heterogeneity as discrete (e.g., developed, hard development). Importantly, the settings variables include a broad but parsimonious suite of attributes that can be used to define the ecological system at any point in time; they are considered primary determinants of ecosystem composition, structure and function, and determine the ecological similarity between any two locations. As such, they play a key role in the ecological integrity assessment, they are used in species' habitat models to represent important habitat components, as appropriate, and are sometimes used in other model components. The settings provide a rich, multivariate representation of important landscape attributes.

1. **All ecological settings data** [updated 7/18/2018]

   https://scholarworks.umass.edu/data/10

2. **Aquatic barriers** [updated 4/20/2018]

   https://scholarworks.umass.edu/data/6

3. **Biomass** [updated 4/24/2018]

   https://scholarworks.umass.edu/data/8

4. **CaCO3** content [updated 4/20/2018]

   https://scholarworks.umass.edu/data/7

5. **Development** (includes development and hard development settings): [updated 4/20/2018]

   https://scholarworks.umass.edu/data/9

6. **Stream gradient** [updated 4/20/2018]

   https://scholarworks.umass.edu/data/11

7. **Traffic rate** [updated 4/20/2018]

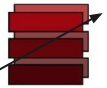
   https://scholarworks.umass.edu/data/12

8. **Temperature** (includes mean annual temperature, growing season degree days, heat index, minimum winter temperature, and maximum summer temperature settings) [updated 6/27/2018]

   https://scholarworks.umass.edu/data/13

9. **Precipitation** (includes total annual precipitation and growing season precipitation settings) [updated 7/17/2018]

Data and Datasets



Designing Sustainable Landscapes: All Ecological Settings

Kevin McGarigal, *University of Massachusetts Amherst*
Brad Compton, *University of Massachusetts Amherst*
Ethan Plunkett, *University of Massachusetts Amherst*
Bill DeLuca, *University of Massachusetts Amherst*
Joanna Grand, *University of Massachusetts Amherst*

Download Data (24768.4 MB) Download Data

Publication Date
2017

Disciplines
Environmental Sciences | Sustainability

Description
The ecological settings products include a broad suite of static as well as dynamic abiotic and biotic variables representing the natural and anthropogenic environment at each location (cell). Static variables are those that do not change over time (e.g., elevation, incident solar radiation). Dynamic settings are available for 2010 and 2080; static settings are available for 2010. Dynamic variables are those that change over time in response to succession and the drivers (e.g., growing season degree days, traffic rate). Most of the settings variables are continuous and thus represent landscape heterogeneity as continuous (e.g., slope, biomass), although some are categorical and thus represent heterogeneity as discrete (e.g.,

II.
What About the Policy?

Impetus for Developing A Data Repository Policy

Request for data deposit for sensitive data:

- Included children's names, ages, "I don't know who these kids are, so no one could find out"
- Expectation from faculty that we would review submissions for private information and anonymize

Realization: We needed to have policies & information to point folks towards!

Bonus points: Written policy would enable us to turn on self-submissions

Initial Decisions

The Data Services Librarian and OA & IR Librarian would work together with the rest of the Data Working Group to collaboratively write the policy

Policy would be based on real experiences

Policy could be added to at any time

Emphasis on flexibility and responsiveness

Writing the Policy

Don’t reinvent the wheel:

- Surveyed policies from existing data repositories

Create outline:

- Each member of the Data Working Group took a section (or two)

Writing the Policy (Cont'd)

Timeline for research and writing: **one month**

Collaborative review for tone, flow, standardization of terms

Post policy publicly with attribution:

Policies used in the creation of this document:

General Statement	DRUM
Data Collection Policy	DRUM
End-User Access Policy	DRUM, eScholarship@UMMS
Deposit License	DRUM, ScholarWorks
Submission Agreement	DRUM, ScholarWorks
Boilerplate language for grants	DRUM, Florida International University
Registering Data	University of Nebraska-Lincoln

III.
Continued Improvements

Mandatory README Files

README files for Data and Datasets

Download the README.txt template

A README file is a text file that contains information about other files in a directory. README files date back to the 1970s as a way to capture general information about a piece of software, program, utility, or a game.

For research data, README files are helpful at providing contextual information to help ensure that the data can be correctly interpreted by yourself at a later date, and by others.

At ScholarWorks@UMass, a README file is required before your data is approved for posting in the repository. The general README file we suggest you complete is available here: https://scholarworks.umass.edu/data/readme.txt. If your discipline already requires more rigorous and detailed README files or metadata, please upload that file with the data you are depositing.

Examples

Example README files are available here:

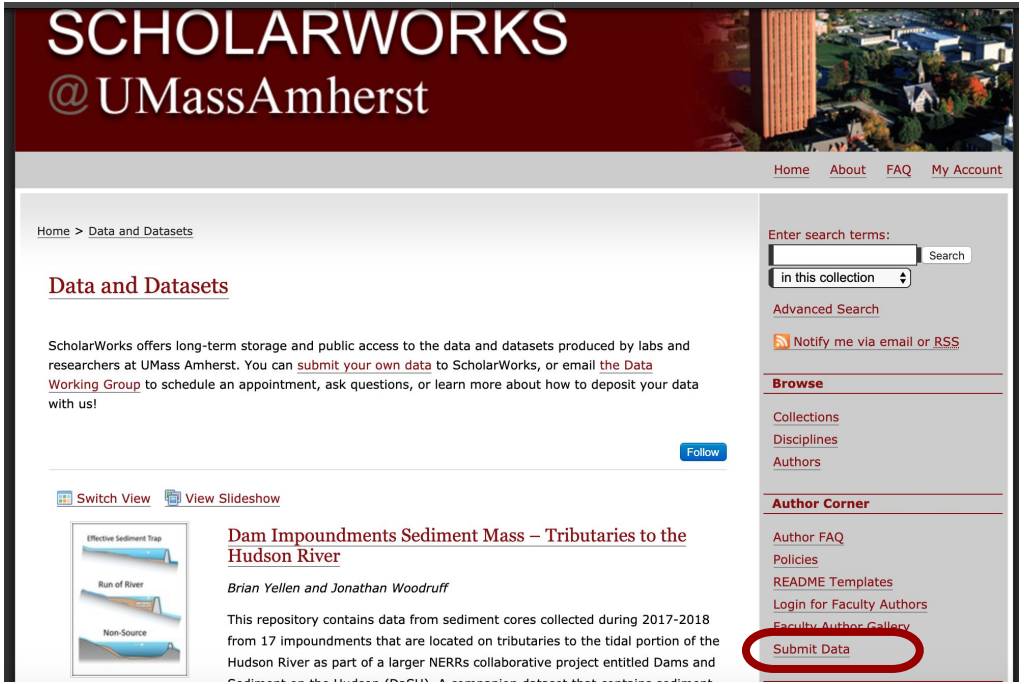
- Example 1
- Example 2
- Example 3

If you have questions about completing a README file, or other questions about sharing your data, please contact the Data Working Group at dwg@library.umass.edu.

```
This README.txt file was generated on <YYYYMMDD> by <Name>


-------------------
GENERAL INFORMATION
-------------------


1. Title of Dataset:




#
# Authors: Include contact information for at least the
# first author and corresponding author (if not the same),
# specifically email address, phone number, and institution.
# Contact information for all authors is preferred.
#

2. Author Information
<create a new entry for each additional author>

  First Author Contact Information
        Name:
           Institution:
           Address:
           Email:


  Corresponding Author Contact Information
        Name:
           Institution:
           Address:
           Email:


  Author Contact Information (if applicable)
           Name:
           Institution:
           Address:
           Email:



----------------------
DATA & FILE OVERVIEW
----------------------

#
# Directory of Files in Dataset: List and define the different
# files included in the dataset. This serves as its table of
# contents.
#
```

Self-Submission vs. Mediated Submissions

Majority of submissions are still mediated, but we have turned on the self-submit link.



Current Stats (Shameless Self-Promotion?)

Total: 112 posted datasets

Downloads (October 2017 - April 2020): 2,284

Most downloaded dataset: “Ionoelastomer Junctions Between Polymer Networks of Fixed Anions and Cations” (123 downloads since November 2019)

Thanks!

Contact us!

Erin Jerome, ewjerome@library.umass.edu

Data Working Group, dwg@library.umass.edu